Scientific Journal Impact Factor (SJIF): 4.364
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

# PRODUCT ASPECT RANKING BASED ON ONLINE CUSTOMER REVIEWS

**Vaibhav Anil Songire[1] and R. V. Patil[2]**
[1]*M.E. Student (CSE), SSVPS's B. S. Deore, College of Engg. Dhule*
[2]*Asst. Prof. (CSE), SSVPS's B. S. Deore, College of Engg. Dhule*

**Abstract**— Now day's e-commerce is rapidly growing which gives facility for customers to purchase products online. Varieties of brands and lots of products have been offered online and numbers of customer reviews are available on internet. These reviews are important for the customers as well as for the sellers. Most of the reviews are disorganized so it generates difficulty for using valuable information. The product aspect ranking system, which identifies important aspects of products from online customer reviews and improve the usability of different reviews. The important product aspects are recognized using two observations: - 1) the important aspects are mostly commented by a large number of users 2) customers opinions on the important aspects are greatly effect on the overall opinions of the product. This paper shows the experimental results on review set of four popular products in two domains.

**Keywords**— Customer reviews, product aspects, aspect identification, sentiment classification, aspect ranking

## I. INTRODUCTION

In the recent years peoples trend towards online shopping increases day by day. There is a rapid expansion in e-commerce. There are many online retail shopping sites available and they indexed millions of products for selling. For example, Amazon.com has collection of more than 36 million products. Most retail websites encourage customers to write reviews to express their opinions on different *aspects* of the product. Here, an *aspect*, also called *feature*, means an attribute of a particular product. An example of review is "*The battery life of Samsung S6 is amazing.*" shows positive opinion on the aspect "*battery life*" of product *Samsung S6*.

Other than retail websites, there are many forum websites also available. They provide a platform for customers to post reviews on millions of products. For example, CNet.com contains more than seven million product reviews, whereas Pricegrabber.com contains millions of reviews on more than 32 million products in 20 categories over 11,000 sellers [3]. Such numerous customer reviews contain rich and valuable knowledge and have become an important source for both customers and sellers. Customers commonly search quality information from online reviews for purchasing a product, while many sellers use online reviews as important feedbacks in their product development, marketing, and customer relationship management strategies [2].

Generally, a product may have number of aspects. For example, *Samsung S6*has more than hundred aspects such as "*memory*", "*screen size*", "*sound quality*", "*camera*." Some aspects are most important than the others, and have more impact on the customers decision making as well as seller's product development schemes. For example, aspects of *Samsung S6*, e.g., "*memory*" and "*camera*," are considered important by most of the customers, and are most important than the others such as "*color*" and "*buttons*". Hence, identification of important product aspects plays an important role to improve the usability of numerous reviews and it is beneficial to both customers and sellers [1].

Customers can easily make purchasing decision by giving attention to the important aspects, while sellers can focus on the enhancement of product quality. However, manual identification of important aspects from huge number of reviews is impractical. Therefore, an approach product aspect ranking is proposed to automatically identify the important aspects from online customer reviews.

## II. LITERATURE SURVEY

This section shows previous work on product aspect ranking system, starting with the product aspect identification with sentiment classification. Existing product aspect identification technique can be classified into two main approaches: supervised and unsupervised.

Supervised learning technique learns an extraction model which is also called as aspect extractor, that aspect extractor is then used to identify aspects in new reviews. For this task Hidden Markov Models, Conditional Random Fields, Maximum Entropy and Naive Bayes Classifier approaches have been used. Wong and Lam used a supervised learning technique to train an aspect extractor. They learned aspect extractor using Hidden Markov Model and Conditional Random Field. All supervised techniques are reasonably effective, but preparation of training examples is time consuming.

In contrast, unsupervised approaches automatically extract product aspects from customer reviews without using training examples. Hu and Liu's works focuses on association rule mining based on the Apriori algorithm to mine frequent item sets as explicit product aspects. In association rule mining, the algorithm does not consider the position of the words in the sentence. In order to remove incorrect frequent aspects, two types of pruning principles were used: compactness and redundancy pruning. The technique is effective which does not require the use of training examples or predefined sets of domain-independent extraction patterns. However, it suffers from two main shortcomings. First, frequent aspects discovered by the mining algorithm might not be product aspects. The compactness and redundancy pruning rules are not able to eliminate these false aspects. Second, even if a frequent aspect is a product aspect, customers may not be expressing any subjective opinion about it in their reviews.

Wu et al also used the unsupervised method. They used the phrase dependency parser to extract noun and noun phrases and then they used a language model to filter out the unwanted aspects. This language model was used to predict the related score of candidate aspects and was built on product reviews. Candidate having low score were filtered out. However, this language model might be biased to frequent terms in the reviews and cannot predict the aspect score exactly as a result cannot filter out noise very efficiently. Then, Popescu and Etzioni developed the *OPINE* system, which extracts aspects based on the *KnowItAll* web information extraction system.

After identification of the important aspects next step is sentiment classification which is used to determine the orientation of sentiment on each aspect. Aspect sentiment classification can be done by using two approaches unsupervised approach and supervised learning approach. Lexicon based approach is typically unsupervised. Lexicon consists of list of sentiment words, which may be positive or negative. This method usually employs a bootstrap strategy to generate high quality lexicon. Hu and Liu have used this lexicon based method. They obtained the sentimental lexicon by using synonym/antonym relation describe in WordNet to bootstrap the seed word set [12].

Hu's method is improved by Ding et al by addressing two issues: opinion of sentiment word would be content sensitive and conflict in review. They derived the lexicon by using some constraints [13].

Second approach is supervised learning approach which classifies opinions on aspects by using sentiment classifier. Sentiment classifier is learned from training corpus which is used to classify the new aspects opinions. Many learning models are applicable for this purpose. Bopong and Lee used three machine learning techniques SVM, Naive Bayes and Maximum Entropy for determining whether the review is positive or negative [10].

There is no previous work study in the product aspect ranking. The product aspect ranking is to predict the ratings on individual aspects. Wang *et al* developed a latent aspect rating analysis model, which aims to determine reviewer's latent opinions on each aspect and the relative emphasis on different aspects. This work concentrates on aspect-level opinion estimation and reviewer rating performance analysis, not on aspect ranking. Snyder and Barzilay expressed a multiple aspect ranking problem. However, the ranking is actually to predict the ratings on individual aspects [1].

*International Journal of Modern Trends in Engineering and Research (IJMTER)*
*Volume 04, Issue 1, [January– 2017]* **ISSN (Online):2349–9745; ISSN (Print):2393-8161**

## III. METHODOLOGY

The proposed product aspect ranking method consists of three components: - 1) Aspect Identification 2) Sentiment Classification on aspects 3) Product aspect Ranking. By giving customer reviews of a product, it firstly identify the aspects from the reviews and then analyze these reviews to find customer opinions on the aspects by using a sentiment classifier and finally rank the product aspects based on importance of aspect by considering aspect occurrence and customers opinions given to each aspect over their whole opinions.

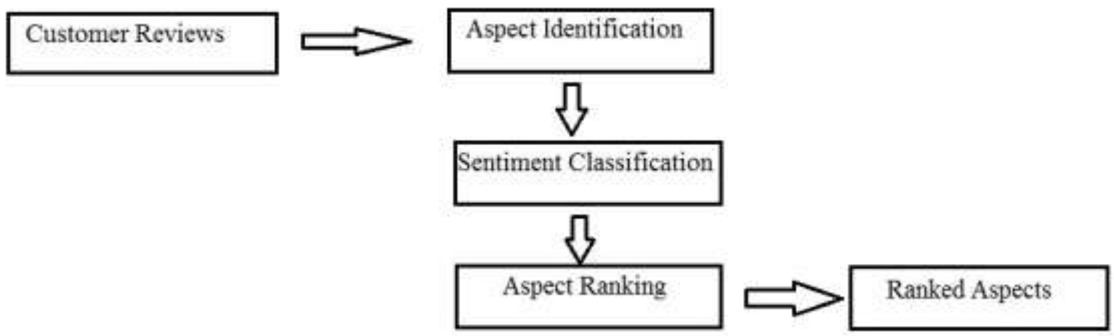


***Figure 1. System Architecture***

### 1. Aspect Identification:-

Customer reviews are composed in different formats on various forum Websites. Several websites enable user to review about the product enable customers to give an overall score on the product, describe brief positive and negative opinions on some product aspects, as well as to enable a review to write a paragraph of detailed review in free text format. Some websites give facility for an overall rating and a section of free-text review. The other websites just need an overall rating and some brief positive and negative opinions on some aspects.

Here we use review set which is in free text format. Firstly divides the free text reviews into sentences, and parses each sentence using parser. The frequent noun terms are then extracted from the sentence parsing trees as product aspects [10]. As the identified aspects may have some synonym terms, for example someone says that "*The battery life of Samsung S6 is amazing*." and someone writes that "*The battery consumption of Samsung S6 is very good.*" so here "*battery life*" and "*battery consumption*" both are same meaning. Then perform synonym clustering to obtain unique aspects. In particular, collect the synonym terms of the aspects as features. We also use the term Important Aspects to perform next actions. Important aspects are those which are mostly commented by large no. of users, so the aspects which have maximum occurrence are the important aspects. The product aspect is the major factor, based on this a customer can purchase product and express their positive opinions as well as negative opinions related to the product, hence customer opinions are key for the seller to develop the business in various fields [1].

### 2. Sentiment Classification on Aspects:-

Sentiment classification is used to extract sentiment (Opinion) from reviews posted by users on products either they are like it or not. Users normally express their opinions as positive, negative, so in sentiment classification reviews are separated on positive and negative basis. The sentiments are classified based on the sentiment words by using sentiment classifier. By giving free text review, it firstly locates the opinionated expression. Generally an opinionated expression is associated with the aspect. Then it gives that opinion in the account of that aspect. For example, "*Samsung S6 has good display*". Here the opinion they have used is good for the aspect display, which is positive sentiment, here the user is expressing their emotions about mobile display which is positive [6].

**3. Product Aspect Ranking:-**

After sentiment classification next step is aspect ranking. Aspect based Ranking is used to identify important aspect of product from reviews given by the users. The important aspects are commented again and again in customer review and the customers opinions on the important aspects are greatly impact their whole opinions on the product [7].

Here $\mathcal{R} = \{r_1, \ldots, r_{|\mathcal{R}|}\}$ denote a set of customer reviews of a particular product. In each review $r \in \mathcal{R}$ customer expresses the opinions on multiple aspects of a product, and finally assigns an overall rating $O_r$. Suppose there are *m* aspects $A = \{a_1, \ldots, a_m\}$ in the review corpus $\mathcal{R}$ totally where $a_k$ is the *k*-th aspect. Customer opinion on aspect $a_k$ in review r denoted as $O_{rk}$.The overall rating $O_r$ is generated based on a weighted aggregation of the opinions on specific aspects, as $\sum_{k=1}^{m} \omega_{rk} o_{rk}$ where each weight $\omega_{rk}$ measures the importance aspects $a_k$ in review r.$o_{rk}$ is the opinion on aspect $a_k$. In matrix form as $\boldsymbol{\omega_r}^T \boldsymbol{o_r}$ where $\boldsymbol{\omega_r}$ denotes a vector of weights and $\boldsymbol{o_r}$ is the opinion vector with each dimension indicating the opinion on a particular aspect. Specifically, the observed overall ratings are assumed to be generated from a *Gaussian Distribution*, with mean $\boldsymbol{\omega_r}^T \boldsymbol{o_r}$ and variance $\sigma^2$ as:-

$$\mathcal{P}(O_r) = \frac{1}{\sqrt{2\Pi\sigma^2}} exp\left\{-\frac{O_r - \omega_r{}^T o_r}{2\sigma^2}\right\}$$

In order to take the uncertainty of $\boldsymbol{\omega_r}$ into consideration assume $\boldsymbol{\omega_r}$ as a sample drawn from a *Multivariate Gaussian Distribution* as:

$$\mathcal{P}(\omega_r) = \frac{1}{(2\Pi)^{m/2}|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(\omega_r - \mu)^T \Sigma^{-1}(\omega_r - \mu)\right\}$$

where $\boldsymbol{\mu}$ and $\Sigma$ are the mean vector and covariance matrix, respectively.

**Algorithm :-**

**Input** : Consumer review corpus $\mathcal{R}$, each review $r \in \mathcal{R}$ is associated with an overall rating $O_r$, and a vector of opinions $o_r$ on specific aspects.

**Output** : Importance scores $\varpi_k|_{k=1}^{m}$ for all the m aspects.

1. ***While** not converged **do***
2. *Update*$\{\omega_r\}_{r=1}^{|\mathcal{R}|}$
3. *Update*$\{\mu, \Sigma, \sigma^2\}$
4. ***End while***
5. *Compute aspect importance scores*$\{\omega_k\}_{k=1}^{m}$

Where $\{\omega_r\}_{r=1}^{|\mathcal{R}|}$ are the importance weights and $\{\mu, \Sigma, \sigma^2\}$ are the model parameters. While $\{\mu, \Sigma, \sigma^2\}$ be estimated from review corpus $\mathcal{R} = \{r_1, \ldots, r_{|\mathcal{R}|}\}$ using the maximum likely hood(ML) estimation, $\boldsymbol{\omega_r}$ review *r* can be optimized through the maximum a posteriori (MAP) estimation. Since $\boldsymbol{\omega_r}$ and $\{\mu, \Sigma, \sigma^2\}$ are coupled with each other. They iteratively optimize $\{\omega_r\}_{r=1}^{|\mathcal{R}|}$ and $\{\mu, \Sigma, \sigma^2\}$ in alternating optimization technique.

After obtaining the importance weights $\boldsymbol{\omega_r}$ for each review $r \in \mathcal{R}$ they compute the overall importance $\varpi_k$ and $a_k$ by integrating its importance scores over the reviews as $\varpi_k = (\Sigma_{r\in\mathcal{R}} \omega_{rk})/|\mathcal{R}_k|$ where $\mathcal{R}_k$ is the set of reviews containing $a_k$. According to $\varpi_k$ the important product aspects can be identified [1].

## IV. DATASET

Here we use the dataset of four products in two different domains. Two mobiles which are Nokia 6600 and Nokia 6610 and other two are routers which are Linksys and Hitachi. Dataset contains customer reviews in free text format.

 

## V. PERFORMANCE EVALUATION

To evaluate the performance of aspect ranking, the mostly used method is Normalized Discounted Cumulative Gain at top k as evaluation metric. By giving ranking list of aspects it is calculated as,

$$NDCG@k = \sum_{i=1}^{k} \frac{2^{t(i)} - 1}{\log(1 + i)}$$

Where, i is the rank of particular aspect, t(i) is the importance degree of the aspect at position i. k is the no. of aspects of the product. Importance degree t(i) judged by three important levels i.e. “Un-important”=1, “Ordinary”=2, and “Important”=3.

In order to evaluate the effectiveness of aspect ranking the aspect ranking algorithm is compared with frequency based method. Frequency based method ranks the aspects according to aspect frequency. It only captures the aspect occurrence information and neglects to consider the opinions on specific aspects [1].

## VI. RESULTS

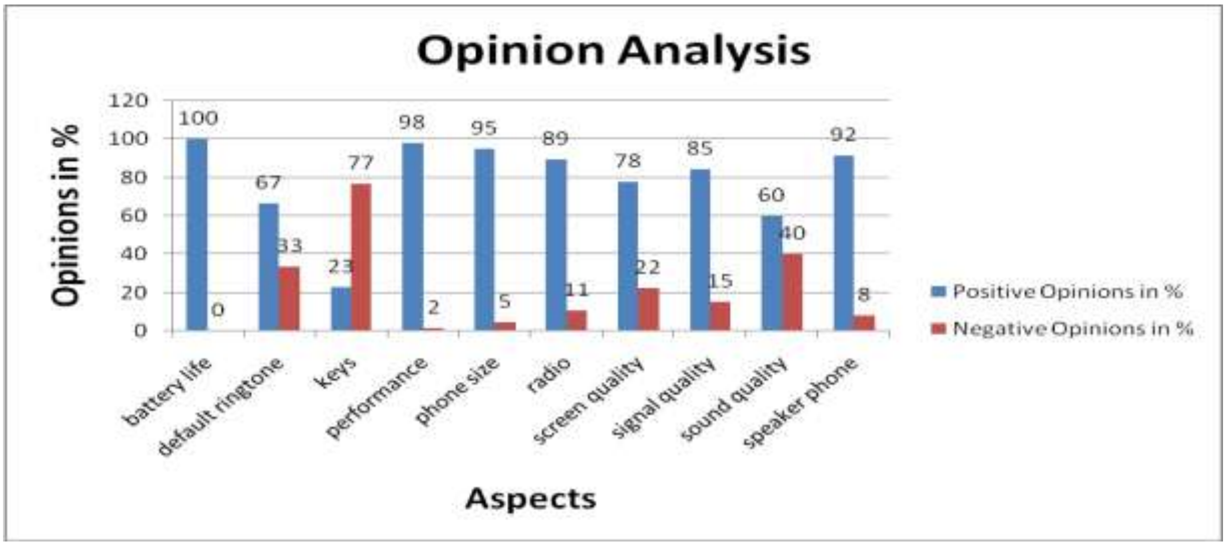


*Figure 2. Sentiment analysis of aspects for product Nokia 6610*

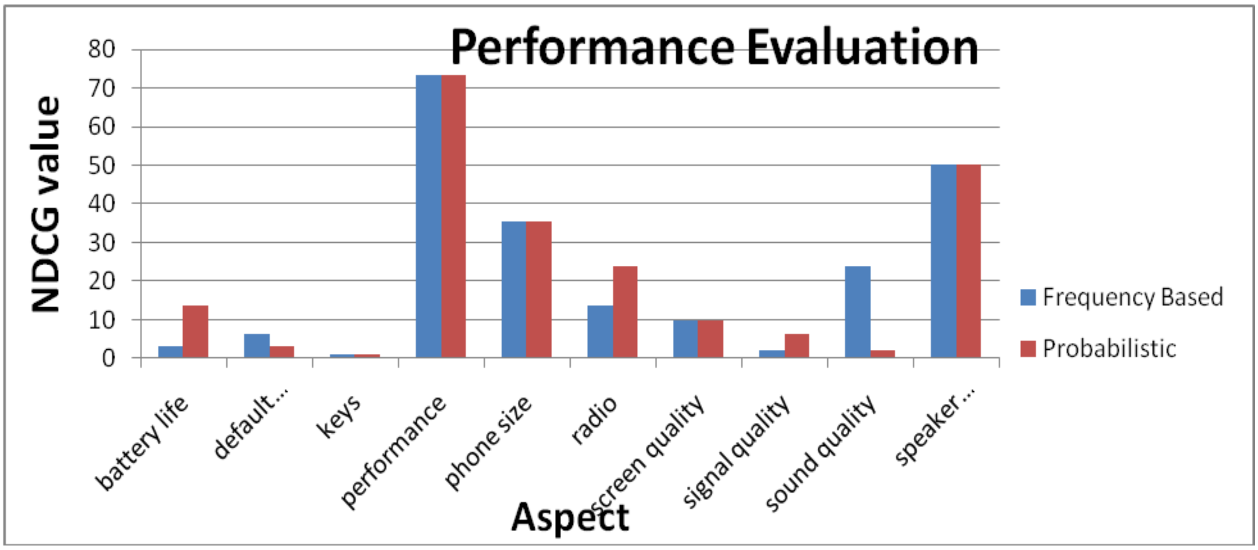


*Figure 3. Performance evaluation using NDCG for product Nokia 6610*

 

## VII. CONCLUSION

Product aspect ranking system contains three main steps i.e. product aspect identification, sentiment classification and aspect ranking. It firstly finds out aspects and their opinions from free text reviews. Then product aspect ranking algorithm explores aspect frequency and opinions and then aspects are ranked according to final importance score. We have used dataset of four products in two domains. An experimental result shows the effectiveness of product aspect ranking system.

## REFERENCES

[1] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang,Meng Wang, Tat-Seng Chua,“Product Aspect Ranking and Its Applications” IEEE Transactions On Knowledge And Data Engineering, vol. 26, no. 5, May 2014, pp. 1211-1224
[2] ShahurajPatil, Jyoti Raghatwan , “An Efficient Product Aspect Ranking and its Application: A Review”, International Journal of Science and Research, vol. 3 , no.12, December 2014, pp.2048-2051
[3] Mr. S. P. Ghode, Prof. S. S. Bere, “Exploiting and Ranking Dominating Product Features through Communal Sentiments”, International Journal on Recent and Innovation Trends in Computing and Communication, vol. 3, no. 6, June 2015, pp. 3894 – 3900
[4] Mr. A. S. Kamale, Mr. S. P. Ghode, Prof. P. B. Dhainje, Mr. A. V. Moholkar, “A Survey on Feature-Sentiment Classification Techniques”, International Journal on Recent and Innovation Trends in Computing and Communication , vol. 2, no.12, December 2014, pp. 3972 – 3978
[5] CharushilaPatil, Prof. P. M. Chawan, PriyamvadaChauhan, SonaliWankhede, “A Survey on Product Aspect Ranking”,International Journal of Innovative Research in Science, Engineering and Technology, vol. 4, no. 12, December 2015, pp. 12196-12204
[6] Aarati Mahadik, Asha Bharambe, “Aspect Based Opinion Mining and Ranking: Survey”, International Journal of Current Engineering and Technology, vol. 5, no. 6, December 2015, pp. 3589-3592
[7] R. Suganya, " Identifying and Ranking Product Aspects based on Consumer reviews," International Journal for Research in Applied Science & Engineering Technology, vol. 3, no. 1, January 2015, pp. 290-293
[8] A. M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in Proc. HLT/EMNLP, Vancouver, BC, Canada, 2005, pp. 339–346.
[9] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. SIGKDD, Seattle, WA, USA, 2004, pp. 168–177
[10] B. Liu, “Sentiment analysis and subjectivity,” in Handbook of Natural Language Processing, New York, NY, USA: Marcel Dekker, Inc., 2009.
[11] B. Liu, “Sentiment Analysis and Opinion Mining”. Mogarn & Claypool Publishers, San Rafael, CA, USA, 2012.
[12] B. Ohana and B. Tierney, “Sentiment classification of reviews using SentiWord Net,” in Proc. IT&T Conf., Dublin, Ireland, 2009.
[13] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in Proc. WSDM, New York, NY, USA, 2008, pp. 231–240.
[14] RutujaTikait, Prof. RanjanaBadre, Prof. MayuraKinikar “Product Aspect Identification and Ranking System” International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 4, April 2015,pp. 1127-1131
[15] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment classification using machine learning techniques” in Proc.EMNLP, Philadelphia, PA, USA, 2002, pp. 79–86.
[16] W. Jin and H. H. Ho, “A novel lexicalized HMM-based learning framework for web opinion mining,” in Proc. 26th Annu. ICML, Montreal, QC, Canada, 2009, pp. 465–472.
[17] Vaibhav Anil Songire and R. V. Patil “A SURVEY ON PRODUCT ASPECT RANKING BASED ON ONLINE CUSTOMER REVIEWS” International Journal of Modern Trends in Engineering and Research (IJMTER)Volume 03, Issue 02, [February – 2016], pp.448-451

 